Determining binding specificities of cell adhesion molecules from *Drosophila* and other related Dipterans

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in Molecular Genetics in the undergraduate colleges of The Ohio State University

by

Leah Anderson

The Ohio State University
April 2018

Project Advisor: Mark Seeger, Department of Molecular Genetics

ABSTRACT

Neurons of both vertebrates and invertebrates exhibit a complex set of cell-to-cell interactions during successful development of the nervous system. Cell adhesion molecules (CAMs) play an important role in mediating many of these specific and stereotyped cell-cell interactions. I am investigating the binding specificities of two CAMs from Dipteran insects: Lachesin (Lac) and Amalgam (Ama). Ama arose as a duplication of Lac early in Dipteran evolution, and both proteins still share extensive sequence similarity. In the fruit fly *Drosohila melanogaster*, Lac is membrane-linked and homophilically binds itself. Ama, which is secreted into the extracellular matrix, has both a homophilic binding property as well as the ability to heterophilically bind another CAM: the transmembrane protein Neurotactin (Nrt). Despite the high level of amino acid sequence similarity between Ama and Lac, the two proteins are unable to bind each other, and Lac does not display any interaction with Nrt. This project is an attempt to identify the precise domain(s) of Lac and Ama that produce these differences in binding specificity. To accomplish this, chimeric constructs of the three immunoglobulin-like domains of Ama and Lac from *D. melanogaster* have been created and cloned into a vector for regulated expression in Schneider 2 (S2) cell lines. The S2 cells are then to be used for aggregation assays, which will reveal binding patterns of the chimeric proteins. Using this approach, a thorough model can be devised for the specific interactions of Lac, Ama, and Nrt in *D. melanogaster*. Additionally, I am utilizing bioinformatic databases to locate and subsequently clone orthologs of Ama and Lac in other Dipteran species. Testing these clones in further aggregation assays will help

develop a better understanding of how the unique binding properties of Ama and Lac have changed over evolutionary time.

INTRODUCTION AND BACKGROUND

Nervous system development in embryogenesis

The nervous system is perhaps the most complex organ system seen in higher animals. Because of this complexity, cells destined to become neural tissue must begin differentiating at very early stages in embryogenesis and travel considerable distances to reach their final developmental destination. An instrumental part of this differentiation and morphogenesis is the interaction and recognition among various cell types within the developing nervous system. The ability of cells to selectively adhere to each other is an essential morphological activity that occurs during development. This ability accounts for cell type segregation, cell rearrangement and migration, and selective fasciculation of axon bundles (de la Escalera et al. 1990). These specific and stereotyped cell:cell interactions are mediated by cell adhesion molecules (CAMs), which are expressed in specific combinations at the appropriate times, leading to successful neural networking.

Fruit flies as a model system

Drosophila is a powerful model for studying neural cell adhesion molecules because of the ability to fix and stain embryos at various stages in development, the easy identification of new cell adhesion molecules through classical genetic screens, and the availability of a well-characterized and annotated genome. Also, insect CAMs involved in neural development are largely conserved through to the vertebrates. Using these organisms as models can help shed light on developmental processes in both

vertebrates and invertebrates, and enhance our understanding of developmental biology across phylogenies.

Genes of interest

One *Drosophila* CAM, known as Neurotactin (Nrt), is a type II transmembrane glycoprotein with a 503 amino acid C-terminal extracellular domain related to the cholinesterases, and a 323 amino acid N-terminal cytoplasmic domain (Barthalay et al. 1990). Nrt first appears in the blastoderm during stage 5 of embryogenesis, and its expression only persists in neural and epithelial cells through the early pupal stages, particularly at points of cell-cell contact. It is not expressed in any adult tissue, however, supporting its proposed role in larval development (de la Escalera et al, 1990). *nrt*$^{-}$ null mutants display mild defects in neural development, but double mutants of *nrt* and some other CAMs act synergistically to produce severe phenotypic defects in axon fasciculation, guidance, and outgrowth (Spiecher et al, 1998). This synergistic effect is seen with many neural CAMs.

Though Nrt is an adhesion molecule, aggregation assays involving *Drosophila* S2 cells expressing Nrt have shown that it is not able to homophilically bind to itself (Barthalay et al, 1990). In order for Nrt expressing cells to adhere to each other, the secreted CAM Amalgam (Ama) must be present in the media as the ligand that binds Nrt (Fremion et al, 2000). It has been shown through protein truncation that the heterophilic adhesive property of Nrt lies within the N-terminal region of the extracellular domain, between His347 and His482 (Darboux et al, 1996). However, while the extracellular domain of Nrt is responsible for its binding to Ama, the cytoplasmic domain

is still crucial for Nrt's ability to promote cell adhesion (Fremion et al, 2000). It is possible that the cytoplasmic region of Nrt interacts with the actin cytoskeleton of the cell in order to successfully carry out cell adhesion (Liebl et al, 2003).

Another CAM, known as Lachesin (Lac), is a member of the immunoglobulin (IG) superfamily of genes, containing three conserved immunoglobulin domains: two C2-type and one V-type. Each domain contains the two regularly spaced cysteines characteristic of most IG domains. Lac comprises 349 amino acids, beginning with a hydrophobic signal peptide at the N-terminus followed by the three immunoglobulin domains and another hydrophobic C-terminal sequence indicative of linkage to the membrane by a glycosyl phosphatidylinositol (GPI) anchor (Karlstrom et al, 1993). This GPI anchor allows Lac to be tethered to the surface of differentiating neural cells in both the central and peripheral nervous system. Many neurogenic cells express Lac early in development, but its expression becomes restricted to those cells in the center of the embryo that are destined to be neuroblasts later in the developmental process. At that point, Lac becomes involved in axonal outgrowth and differentiation of neurogenic cells and plays an important role in cell recognition and adhesion necessary for formation of the nervous system (Karlstrom et al, 1993). Lac is a homophilic cell adhesion molecule, demonstrating the ability to bind itself in bead aggregation assays (Llimargas et al, 2003). The IG domains allow Lac to act as neuronal "glue," and contribute to defining the neural network (Karlstrom et al, 1993). There is high conservation of Lac across species, indicating that its role is vital for development. For example, grasshopper Lac shares 75% identity with *Drosophila* Lac. This indicates that there was high selective

pressure for a significant portion of the protein, and suggests that Lac carries out instrumental functions for the developing organism (Karlstrom et al, 1993).

Ama, another member of the immunoglobulin superfamily, comprises an N-terminal amino acid signal peptide sequence followed by three immunoglobulin-like domains and a short carboxy terminal (Seeger et al, 1988). Similar to Lac, the three IG-like domains of Ama are conserved, two C2-type and one V-type, with sequences that are 22-36% identical (Seeger et al, 1988). While Ama is 60% similar to Lac, it notably lacks the GPI anchor characteristic of Lac, and is instead secreted into the extracellular space. It has been seen to accumulate on the membranes of neural cells of developing embryos, and first appears in stage 8 of embryogenesis, just after the formation of three germ layers (Seeger et al, 1988). *ama*- mutants do not display a severe phenotype; Ama is not believed to be essential to maintain the developing nervous system. It is possible that functional equivalents exist for this protein, similar to the case for Nrt (Seeger et al, 1988). However, *ama*- pupae exhibit some fasciculation defects similar to those found in *nrt*- mutants (Fremion, et al, 2000).

As mentioned before, Ama has the ability to bind the transmembrane protein Nrt. The similar protein expression patterns of Ama and Nrt *in vivo* support their relationship as ligand and receptor. Additionally, Ama does not display its normal accumulation pattern in nrt^{-} mutants, and Ama is able to promote aggregation of Nrt-expressing S2 cells (Fremion et al, 2000). In addition to its heterophilic interaction with Nrt, Ama is able to homophilically bind itself. S2 cells transfected with a transmembrane version of Ama form homotipic aggregates (Liebl et al, 2003).

Interestingly, the majority of genes involved in development came about due to duplication and divergence. Amalgam is an example of this phenomenon. Ama is unusual among other proteins of the IG superfamily in that all three IG domains are contained within one exon. In most IG superfamily genes, each IG domain is encoded by one or two exons. This is evidence that Ama came about as a duplication (Seeger et al, 1988). The closest gene relative to Ama is Lac, which is 38% identical and 58% similar to Ama when accounting for conservative amino acid substitutions (Karlstrom et al, 1993). This is evidence that Ama arose due to a duplication of Lac. The presence of both Lac and Ama in the derived Dipterans but absence of Ama in the more ancestral Dipteran species indicates a duplication event occurred shortly after the origin of Diptera, around 200 million years ago.

Purpose and significance of this work

Despite the high degree of sequence similarity between Lac and Ama and their apparent common ancestry, the functionality of these two proteins has considerably diverged. While Lac is tethered to the cell membrane and mediates cell:cell adhesion via homophilic binding, Ama functions with both a homophilic binding property and a heterophilic interaction with Nrt (Figure 1). Interestingly, Ama and Lac are not able to bind each other, and Lac displays no interaction with Nrt, despite their similar sequences and structure. The goal of this research is to determine where these binding specificities lie. Which IG domain is responsible for the homophilic binding property of each protein? Which domain allows Ama, but not Lac, to bind Nrt?

A useful system to directly observe the interactions of cell adhesion molecules is through S2 cell aggregation assays. S2 cell lines, which are derived from *Drosophila* embryos, normally exist as single cells in suspension. When transfected with DNA that encodes CAMs and manipulated to express these genes, the cells will form multicellular aggregates. With this method, it can be demonstrated which proteins are able to mediate cell adhesion.

To determine how the protein binding specificity of Ama and Lac has diverged, new combinations of these genes can be created, and the activity of the resulting proteins will provide clues about the specificity of each binding domain. This can be done by rearranging the three immunoglobulin domains of Ama and Lac into unique genetic chimeras and expressing these recombined genes in S2 cells. Aggregation assays between cells expressing chimeric proteins and cells expressing wild-type Lac, Ama, or Nrt then give clues about the distinct protein regions responsible for specificity.

In addition to the binding specificities of *Drosophila melanogaster* Lachesin and Amalgam, this research addresses the changes that have occurred in these genes over time across related species. Evolutionarily, Lac and Nrt are ancient genes that exist in all insects. The relatively recent duplication of Lac in Dipteran evolution led to the existence of Amalgam, which is still rapidly evolving. Lac represents the more ancient function and is highly conserved across species, while the binding specificities of Ama are representative of newly diverged protein-protein interactions.

This project also aims to classify how Lachesin has evolved over time using orthologs of Lac and Ama in species related to *D. melanogaster*. Using online bioinformatic databases, I have identified likely Lac and Ama orthologs in over one

hundred other insect species, most within the Dipteran order. Only Dipterans have Amalgam; the higher Dipterans and close relatives in other orders have only Lac, but no Ama. Analysis of multiple sequence alignments between these orthologs reveals the amino acid locations that are most conserved and therefore most important for developmental function.

The continuation of this research will involve cloning these genes and using them in aggregation assays with the wild-type genes from *D. melanogaster.* My hypothesis is that Lac homologs from other species may still be able to adhere to *D. melanogaster* Lac, due to the high levels of sequence conservation. However, it will be interesting to see if Ama homologs from Dipterans are able to bind to *Drosophila* Ama and Nrt, since Ama is a recently diverged and rapidly evolving gene.

Developmental processes are highly regulated and largely conserved in both invertebrate and vertebrate species. This research provides basic functional and bioinformatic analyses of neural cell adhesion molecules, an integral part of development. This analysis may provide insights about the binding characteristics and molecular evolution of developmental genes across phylogenies.

METHODOLOGY

Construction of HA-tagged genetic chimeras via PCR

Genetic constructs of Lachesin, Amalgam, and Nrt in the pMET vector were previously cloned in the Seeger lab via polymerase chain reaction (PCR) and ligation techniques. The 6 chimeric versions of Ama and Lac were also available, both with the GPI and without: L1L2A3, L1A2L3, A1L2L3, L1A2A3, A1L2A3, A1A2L3. For this project, an HA epitope tag was then added to each secreted chimera via PCR. This makes each chimera detectable in a Western blot to confirm protein presence before performing functional assays. The primers used were Lac-f Ama-f, Lac-secHA-r, and AmaHA-r. The following secreted products were all made: L1L2A3HA, L1A2L3HA, A1L2L3HA, L1A2A3HA, A1L2A3HA, A1A2L3HA, AmaHA, and Lac-secHA. A DNA cleanup protocol was done, and the cleaned up PCR products were then combined in an infusion ligation with linearized pMET. Clontech In-Fusion HD cloning kit was used for this infusion ligation. The ligated plasmids with the desired genetic products were then transformed into competent *E.coli*, and bacteria that had taken up the DNA were put under selection using Ampicillin, since the pMET vector bears an Amp resistance marker. Single colonies of *E.coli* were cultured in liquid SB media overnight in a 37 degree shaker to amplify amount of bacteria with the desired DNA, and boiling plasmid minipreps were performed to extract the DNA. Restriction enzyme digests with EcoRI and BamHI were then set up, and each digested sample was run on a DNA gel to test for presence of an insert of the correct size. Samples that tested positive were cleaned up and sent to a sequencing facility to verify the accuracy of the sequence. Once this

was confirmed, E. coli was re-transformed with the miniprep sample, large cultures were grown overnight, and DNA midiprep extractions were performed in order to obtain high levels of each HA-tagged construct in pMET. Finally, UV spectrophotometry was used to determine the concentration of DNA in each midiprep sample. This allowed for exact microgram measurements for S2 cell transfections.

Stable S2 cell transfections and cell culture

Untransfected S2 cell lines were maintained in the lab and passaged in S2 media, (10% FBS, 1%PSA) at 1:10 dilutions every 10 days. Stable transfections were done at around day 5 after passaging. Using the Cellfectin protocol (Cellfectin II reagent – Gibco by Thermo Fisher Scientific), 1.5mL of cells were combined with 2μg Hyg plasmid and 8μg of each sample in pMET. This Hyg plasmid was used as a selectable marker for the cells that took up DNA in the transfection process; addition of the antibiotic Hygromycin B (Invitrogen by Thermo Fisher Scientific; 200ug/mL) allowed only the Hyg-transfected cells to survive. This selection ensured that all cells used in functional assays contained the DNA of interest. Using this selective method, stable transfected lines were created for each secreted and GPI-linked chimera, Lac and Ama (both secreted and GPI-linked), Nrt, and a Hyg-only control. The stable S2 cell lines were then passaged at 1:5 or 1:10 dilutions every 10-14 days, depending on cell concentration.

Western blotting

Expression of each construct was induced using copper sulfate ($CuSO_4$), which allows for transcription by binding the metallothionein promoter in pMET.

SDS-PAGE and Western blotting were used to test for successful expression of protein with anti-Nrt and anti-HA antibodies. Cells expressing Nrt were pelleted, resuspended with water, and combined with 2X sample buffer, then loaded into a polyacrylamide protein gel. Cells expressing the HA-tagged secreted chimeras were pelleted, and the secreted media was combined with sample buffer and loaded into a separate protein gel. Proteins were transferred from the gels to membranes, and the membranes were incubated with primary antibody (1:2000 ratio) overnight. For the secreted media blot a rabbit anti-HA antibody was used; for the Nrt blot, a mouse anti-Nrt was used. On the second day, several washes were done, and then the membranes were incubated with an alkaline phosphatase (AP) conjugate secondary antibody (1:2000) – either goat anti-rabbit (for the HA-tagged proteins) or goat anti-mouse (for Nrt). AP fluorescence was achieved through the BCIP reaction. This diagnostic showed that all secreted chimeric proteins and Nrt were being expressed at a high level in the stably transfected S2 cell lines. Unfortunately, no antibodies were available for the GPI-linked chimeras, so this test was not performed on those cells.

Cell aggregation assays

Assay of secreted chimeras

S2 cells that were stably transfected with each secreted and HA-tagged chimera, Lac-sec, and Ama were induced with copper sulfate (0.7mM) and allowed to produce

protein overnight. The cells were then pelleted by centrifugation using a low speed clinical centrifuge for 5 minutes, and the media containing secreted protein was transferred to a new tube. For each secreted media sample, 3mL of stable Nrt-expressing S2 cells were pelleted, resuspended in the secreted media, and transferred to a 6-well plate. The residual copper sulfate in the secreted media was sufficient to induce expression of the Nrt cells. The plates were then left on a room temperature shaker. Pictures were obtained after 8 hours.

Assay of GPI-linked chimeras

Stably transfected S2 cells containing Lac and Ama-GPI were stained with fluorescent dil by adding dil, incubating overnight, and washing the cells several times with S2 media. 3mL cell aggregation assays were set up between dil-labeled Lac or Ama-GPI with unlabeled cells expressing Lac, Ama-GPI, and each GPI-linked chimera. A hemocytometer was used to count and normalize cell count across the cell lines used. Copper sulfate was added to each aggregation assay, and the plates were left on a room temperature shaker overnight. Pictures were obtained at about 24 hours.

Cell pull-down assays

Stably transfected Nrt-expressing cells and non-transfected cells (as a negative control) were induced with copper sulfate (0.7mM) and allowed to produce Nrt overnight. These cells were then pelleted in microcentrifuge tubes and resuspended in media containing secreted chimeric proteins obtained from S2 cells expressing these constructs (see above: “Cell aggregation assays”). Each sample was allowed to incubate on a tube rotator at 4^{o}C for about 4 hours. After 4 hours, each sample was

pelleted at 2500g for 2 minutes, washed with 1mL serum-free S2 media, pelleted again, and finally resuspended in 50μL 2X sample buffer. Each cell lysate was run in a protein gel (20μL per sample), and a subsequent Western blot was done (using the same parameters as described earlier).

Bioinformatics

Flybase

Genomic, transcriptomic, and proteomic information for *Drosophila* species and other closely related Dipterans are well characterized and annotated on Flybase. DNA, RNA, and protein sequence information for Ama and Lac was obtained from this database.

Basic Local Alignment Search Tool (BLAST)

The BLAST search was utilized on several genomic database platforms, including Flybase, Vectorbase, and the National Center for Biotechnology Information (NCBI) websites. The "tblastn" function (using a protein query to search translated nucleotides) was used to search for *D. melanogaster* Lac and Ama orthologs in over one hundred other insect species, most belonging the order Diptera.

Clustal Omega and MEGA bioinformatics tools

Multiple sequence alignments of orthologous protein sequences were generated using Clustal Omega and the "muscled alignment" tool on MEGA7. The muscled alignment was then used to generate a maximum likelihood tree in MEGA7.

RESULTS AND DISCUSSION

S2 cells provide a useful model to demonstrate CAM binding specificity

It has previously been shown in our lab that cell aggregation assays between stable S2 cell lines are an effective way to demonstrate the interactions of CAMs such as Lac, Ama, and Nrt (Libel et al, 2003). Transfected S2 cell lines exist as cells in suspension, but begin to express the transfected gene upon addition of copper sulfate to the media. When Lac-transfected S2 cells are induced with copper sulfate, large multicellular aggregates form. The same result is seen with Ama-transfected cells. Cells expressing Nrt, however, do not aggregate on their own. They do form aggregates when Ama is present in the media but not when only Lac-sec is in the media (Figure 2). These patterns support the known binding specificities of these CAMs.

This project utilized genetic chimeras of the three IG domains of Lac and Ama generated through PCR (Figure 3). Each unique combination of IG domains 1, 2, and 3 was created, both with GPI anchors and without (so that the protein is secreted). Ama-GPI and Lac-sec were also constructed. An HA epitope tag was added to each secreted construct so that the proteins were easily detectable with a single antibody in a Western blot. The chimeras were then all transfected into S2 cells, and stable cell lines were created via selection for successfully transfected cells by addition of Hygromycin B. Transfected cells usually took several weeks to reach a high abundance of cells, at which point they could be induced and combined in cell aggregation assays.

The first IG domain of Lac and Ama confers homophilic binding

When cells containing Lac, Ama-GPI, or any of the GPI-linked chimeras are induced, large cell aggregates form. To investigate the observed homophilic binding property, mixed aggregation assays were done between cells expressing Lac or Ama-GPI and those expressing every GPI-linked chimera. In each mixed assay, one cell type was labeled with red dil, and the other was unlabeled. The control assay of labeled Lac cells combined with unlabeled Lac cells formed heterotipic aggregates (the aggregates contained both labeled and non-labeled cells). The assay of labeled Lac cells with unlabeled Ama-GPI cells, however, led to the formation of homotipic aggregates, where the aggregates either completely fluoresced or did not at all (Figure 4).

Aggregation assays of dil-labeled Lac combined with unlabeled cells expressing each GPI chimera were set up. Similar to the assay between Lac and Ama-GPI, assays between Lac and chimeras containing the first IG domain of Ama led to the formation of homotipic aggregates, suggesting that Lac was not interacting with these chimeras. Whenever the first IG domain of Lac was present in the chimera, however, those cells formed heterotipic aggregates (Figure 4). This indicates that the first domain is necessary for the homophilic binding specificity seen in Lac.

The same sort of result was seen with Ama-GPI: when labeled cells expressing Ama-GPI were combined with cells expressing chimeras with the IG domain L1, there was no interaction observed between the two cell types. When combined with cells expressing chimeras with A1, heterotipic aggregates formed, supporting the idea that

the first IG domain of Amalgam is necessary for its homophilic binding specificity. This result supports the previous results seen in this lab (Table 1).

The second and third IG domains of Ama are necessary for the Ama:Nrt interaction

Another cell aggregation assay was utilized to determine the binding specificity of Ama to Nrt. Stable S2 cell lines containing Ama, Lac-sec, and each of the six secreted chimeras were induced with copper sulfate and allowed to produce protein overnight. The cells were then pelleted, and the secreted media was then added to both Nrt expressing cells and non-transfected S2 cells. After several hours of incubation on a shaker, aggregation was observed in several of the samples. The Nrt cells that were combined with Ama-containing media formed large multicellular aggregates, but the cells with Lac-sec in the media did not aggregate. As a negative control, the aggregation assays of secreted media combined with non-transfected cells showed no aggregates. The secreted chimeras L1L2A3, L1A2L3, A1L2L3, A1A2L3, and A1L2A3 in solution with Nrt-expressing cells led to no formation of aggregates. The chimera L1A2A3, however, was able to facilitate cell adhesion comparable to that observed with wild-type Ama (Figure 5A). This result suggests that the second and third IG domains of Ama comprise the necessary component of Ama's specific, heterophilic interaction with Nrt. This leads to a possible model of Ama-mediated cell adhesion as shown in Figure 5B.

Another approach used to study the Ama:Nrt interaction was a pull-down binding assay. Induced Nrt-expressing cells were suspended in media containing each secreted protein, and inverted for several hours. This was also done on a set of

untransfected control cells. The cells were then pelleted and washed. Cell lysates were created for each sample; the lysates were then run in a protein gel and detected in a Western blot. Two Western blots were done: one for the control cells and one for the Nrt-expressing cells. On the control blot, protein bands can be seen in each lane, revealing that the secreted proteins were still present in the cell lysate after washing (Figure 6A). On the Nrt blot, however, there is a very strong band for Ama and the L1A2A3 chimera (with much greater intensity than any other band), which supports the cell aggregation assay result (Figure 6B). A high level of background was seen on these Western blots, however, making the results less convincing. The observed background may be due to insufficient washing, random endocytosis, or presence of unknown interacting proteins endogenously expressed by S2 cells. The next approach to be taken on this question is a co-immunoprecipitation experiment, which would ideally produce less background.

The findings from these cell aggregation assays and pull-down assay lead to a possible model where the second and third domains of Ama are bound to the extracellular portion of Nrt, and the first domain is homophilically bound to itself, facilitating adhesion between two Nrt-expressing cells (Figure 5B). However, more research needs to be done to reach a conclusive answer.

Idenification of gene orthologs using bioinformatic databases

In addition to functional analysis, a bioinformatic analysis was conducted using online genomic databases. BLAST searches using the *D. melanogaster* Ama and Lac protein sequences as queries were done to identify the most likely orthologs in 115

other insect species, mostly of the Dipteran order. A subset of species was selected, and the translated orthologs from these species were aligned using the "muscled alignment" tool in MEGA7. From this alignment, a maximum likelihood phylogenetic tree was generated (Figure 6). This tree includes 14 Dipteran species and one outgroup: *Tribolium castaneum* (red flour beetle). Orthologs for both Lac and Ama were identified in all 14 Dipteran species, but only Lac is found in *Tribolium*. This supports the notion that Lac is the more ancestral gene compared to Ama, and that Ama resulted in a duplication of Lac after the emergence of early Dipterans. Another feature of this tree is that it uses varying branch length to represent evolutionary divergence. It can also be seen that the branch lengths for the Lac orthologs are shorter than the branches for the Ama orthologs. This supports the idea that Ama, as a duplicated gene, was under much less selective pressure than Lac, and has had the freedom to more rapidly accumulate differences over evolutionary time.

This bioinformatic analysis provides the framework for the continuation of this project from an evolutionary standpoint. Now that orthologous sequences have been found in other species, PCR can be used to clone these genes for subsequent functional assays like the one described in this thesis. This will provide insight into the divergence in function of these cell adhesion molecules over evolutionary time.

FIGURES

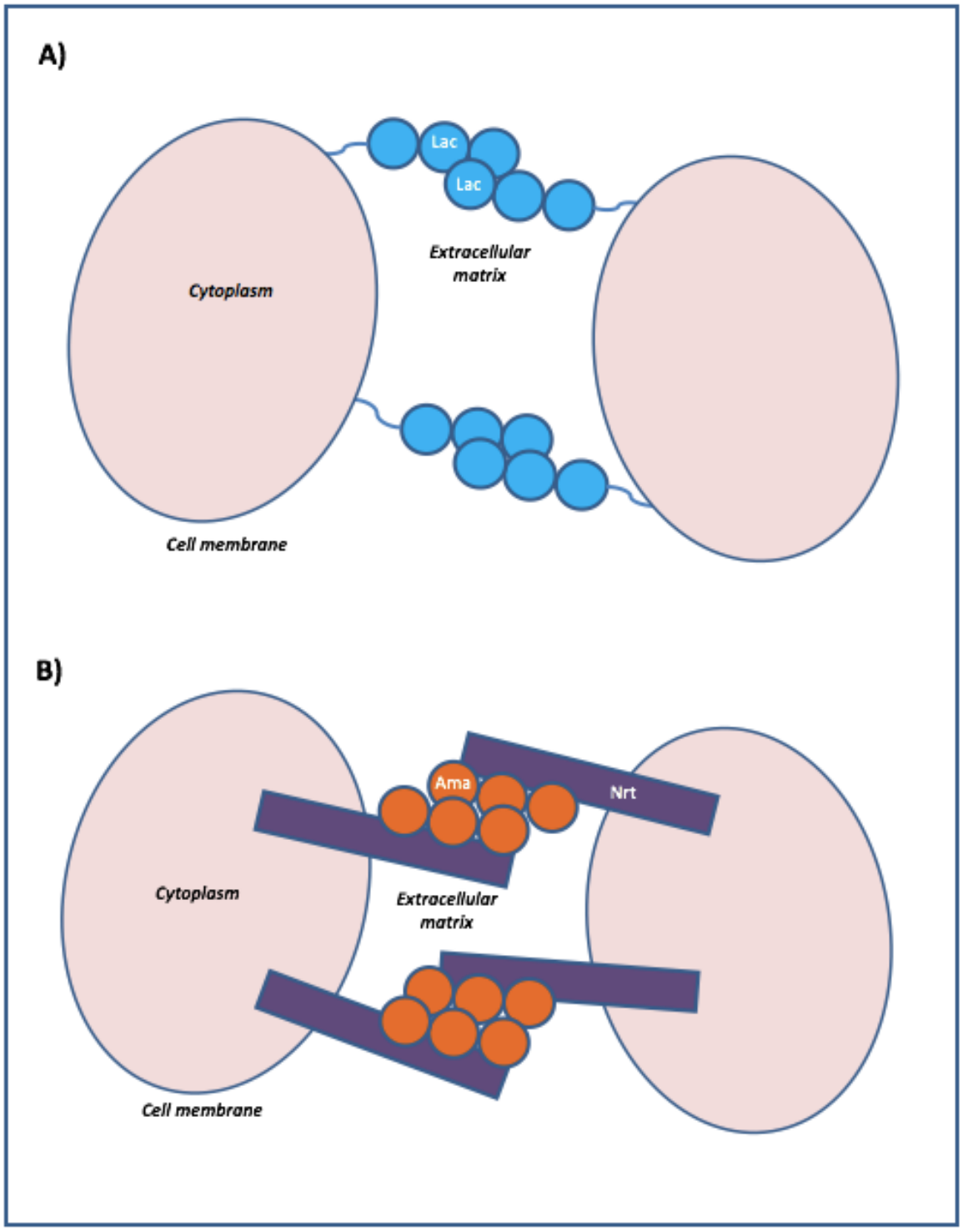


Figure 1. Models of Lac- and Ama-mediated cell adhesion
Models for cell adhesion facilitated by Lac (A), or Ama and Nrt (B). Lac is tethered to the membrane via GPI anchor, Nrt is transmembrane, and Ama is secreted into the extracellular space.

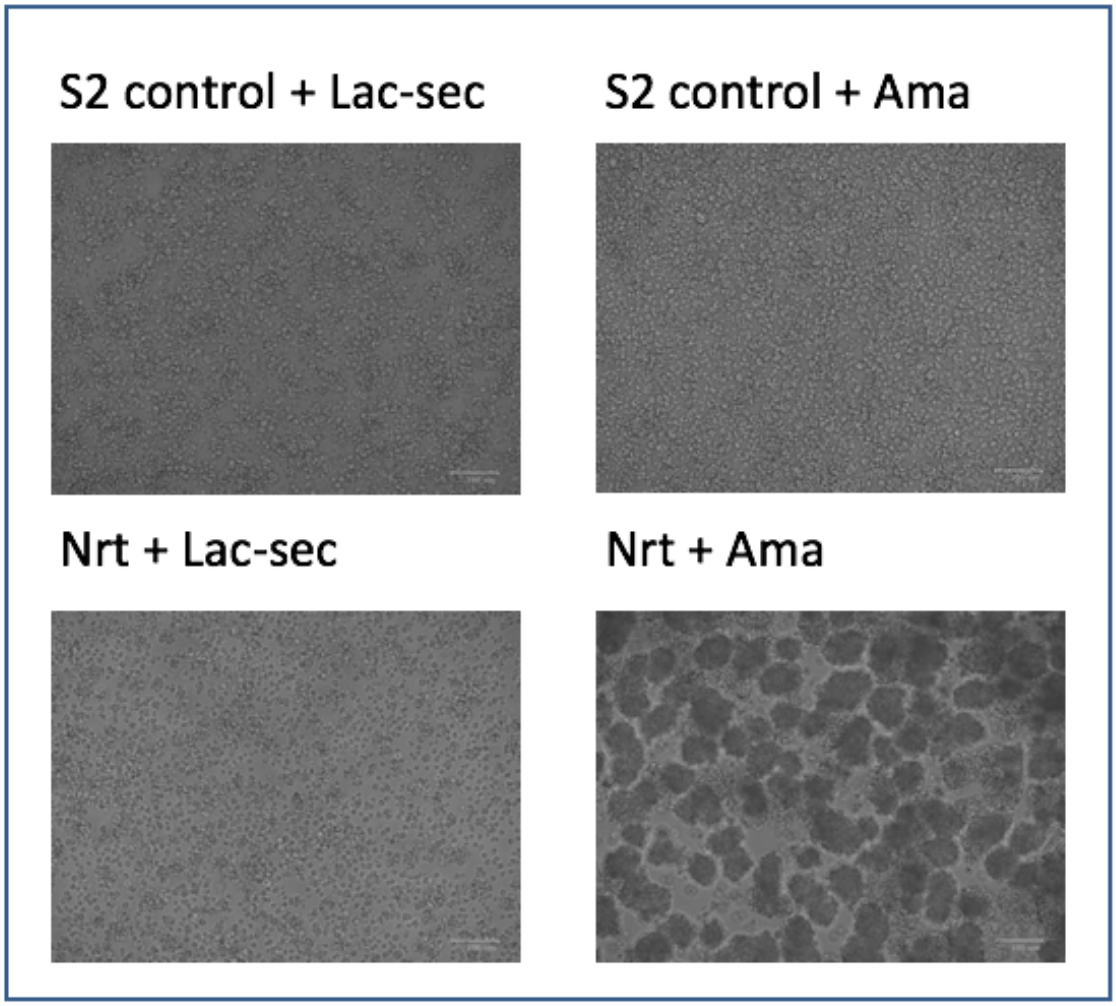


Figure 2. Cell aggregation assays of Nrt-expressing cells and control S2 cells
When induced with copper sulfate and combined with media containing secreted protein, control S2 cells do not aggregate (top panels). Cells expressing Nrt do not aggregate when secreted Lac is present in the media (bottom left), but form large aggregates in the presence of Ama (bottom right).

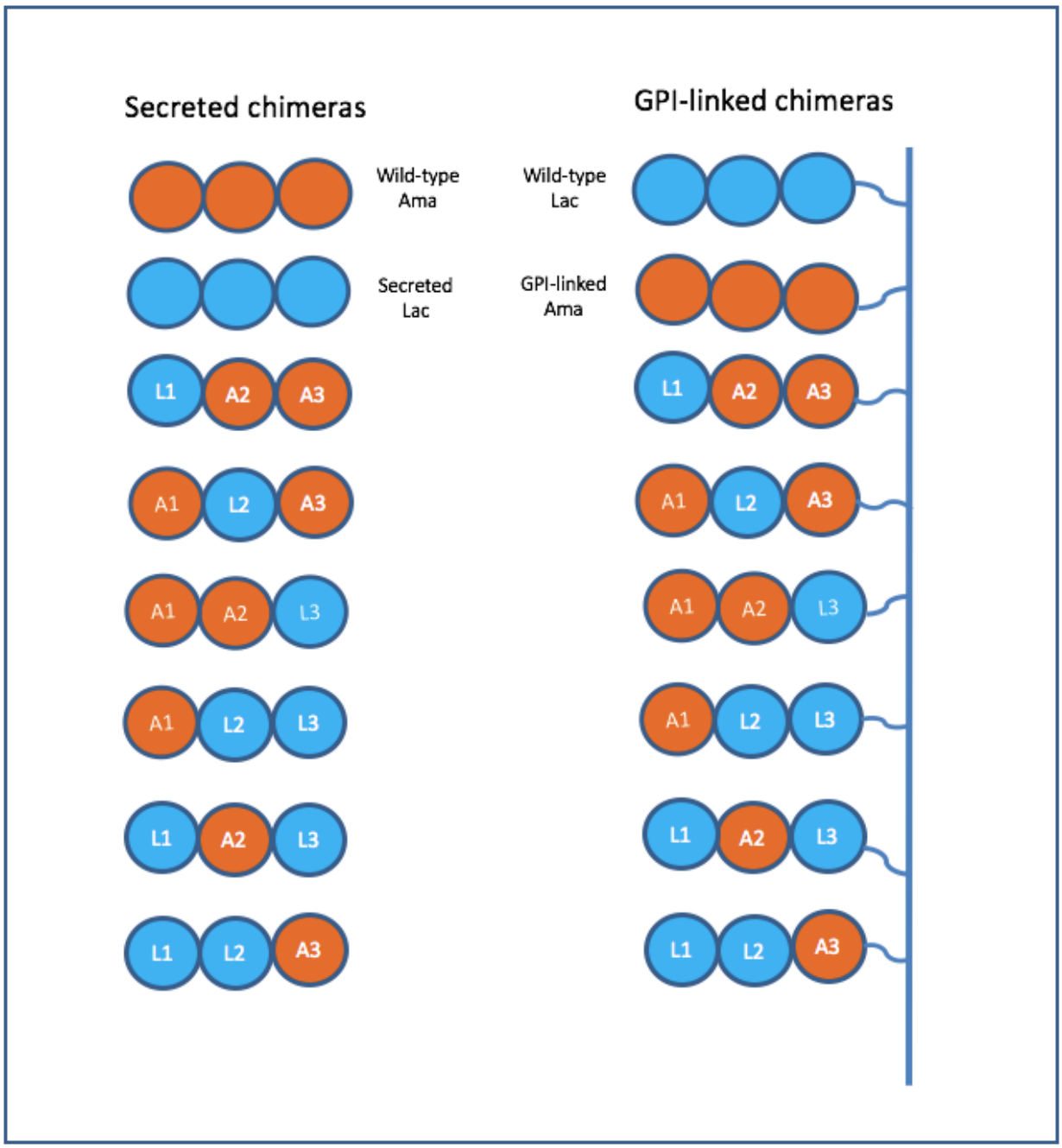


Figure 3. Genetic chimeras
Chimeric constructs of the three IG domains of Lac and Ama. Orange circles represent the IG domains of Ama, and blue circles represent those of Lac. Each possible combination of both secreted and GPI-linked chimeras was generated using PCR. Included in this group are a secreted version of Lac and a GPI-lined version of Ama.

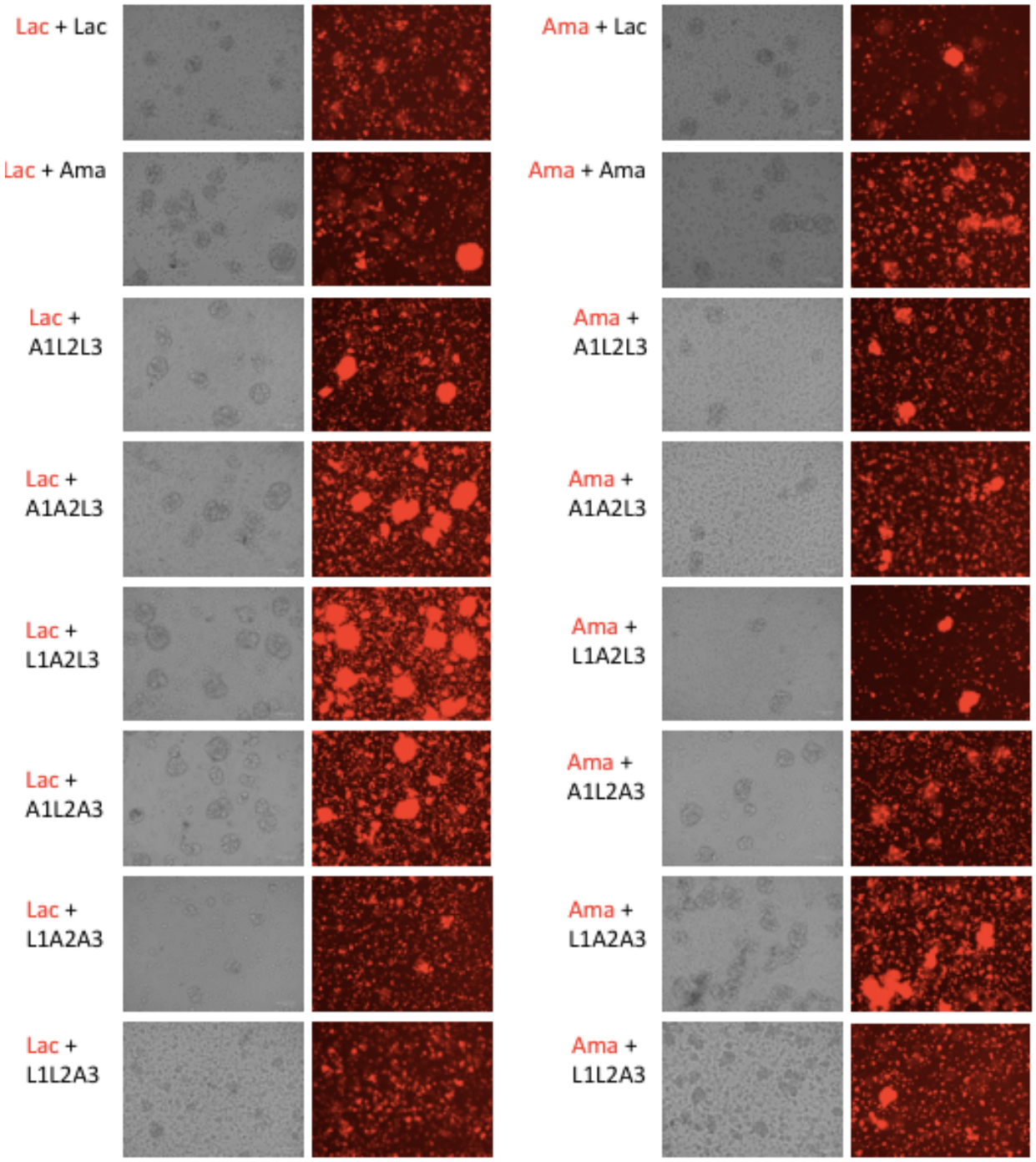


Figure 4. Cell aggregation assays of cells expressing GPI-linked proteins
Labeled cells expressing Lac (left) and Ama (right) were combined with unlabeled cells expressing Lac, Ama-GPI, and each GPI-linked chimera.

	binds Lac	binds Ama
Lac	yes	no
Ama-GPI	no	yes
L1L2A3-GPI	yes	no
L1A2L3-GPI	yes	no
A1L2L3-GPI	no	yes
A1A2L3-GPI	no	yes
A1L2A3-GPI	no	yes
L1A2A3-GPI	yes	no

Table 1. Homophilic binding specificity of Lac and Ama
Cells expressing chimeras with the first domain of Lac or Ama were able to form heterotipic aggregates with cells expressing Lac or Ama-GPI, respectively.

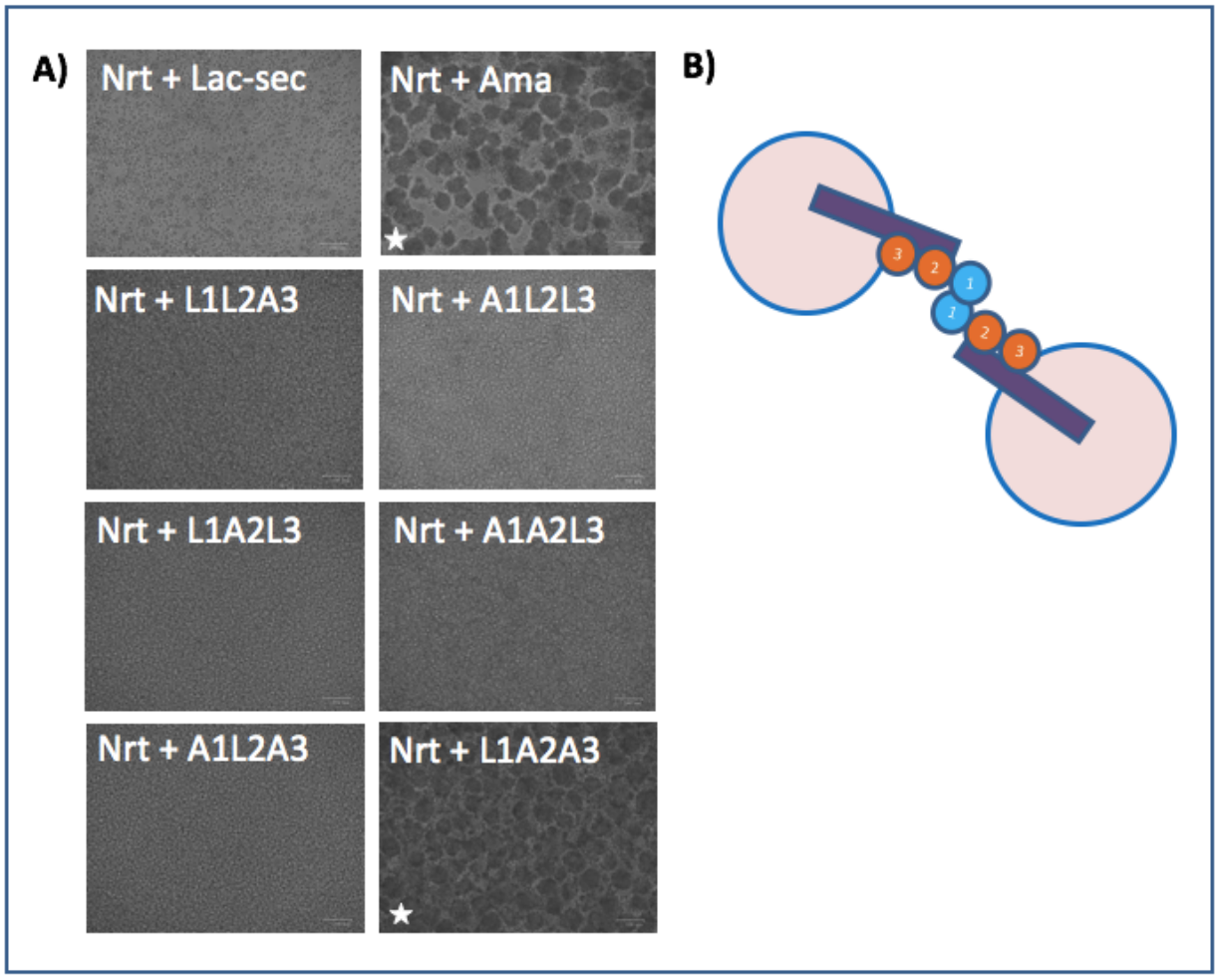


Figure 5. Cell aggregation assays to study the interaction of Ama and Nrt

A) Cell aggregation assays of Nrt-expressing cells in media containing secreted chimeric proteins. Large aggregates are seen when Ama is present in the media. The only chimera able to replicate the function of wild-type Ama is L1A2A3.

B) Possible model for the adhesion between Nrt-expressin cells with secreted L1A2A3 in the media.

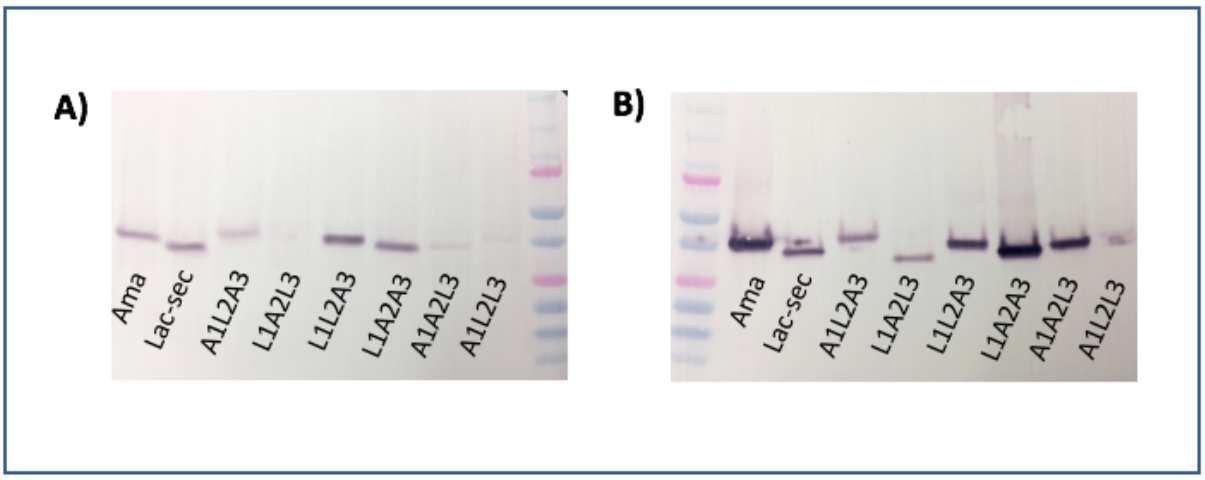


Figure 6. Pull-down assays

A) Control Western blot. Cell lysates of untransfected control cells combined with media containing secreted proteins. Presence of the secreted proteins was detected, showing a high level of background in this experiment.

B) Nrt blot. Cell lysates of Nrt-expressing cells combined with media containing secreted proteins. Background is high, but the protein bands for Ama and L1A2A3 are noticeably brighter than the rest.

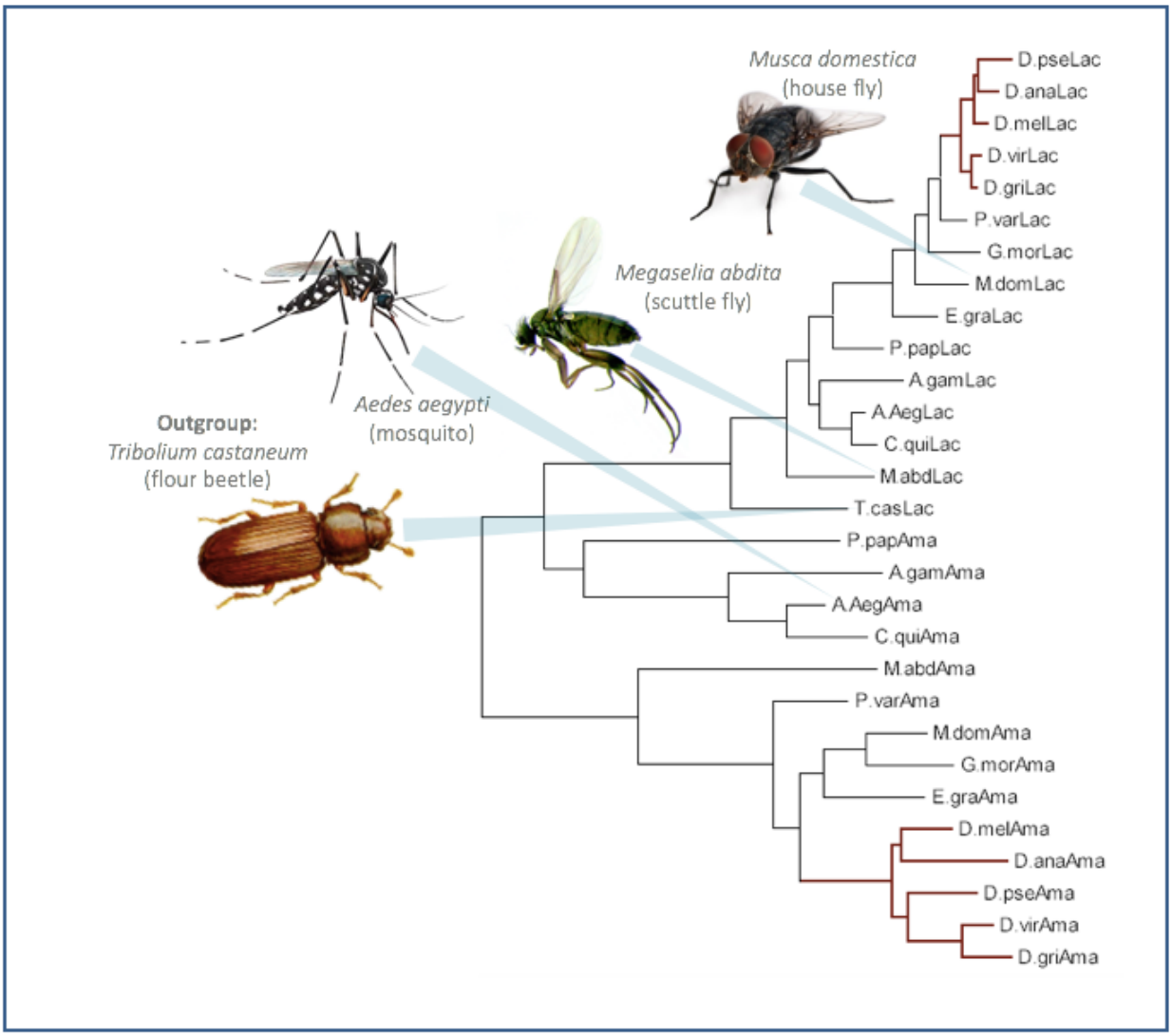
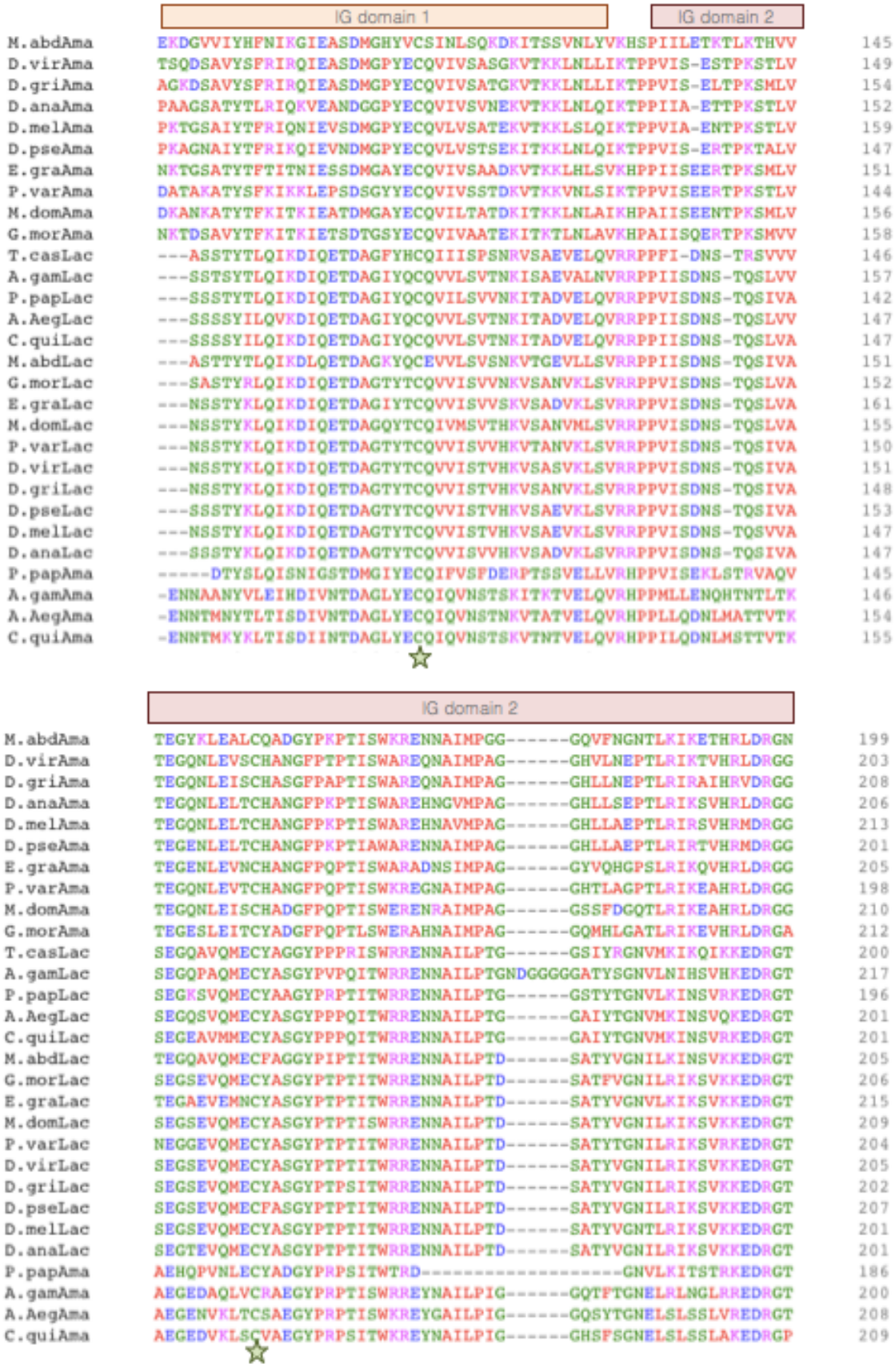
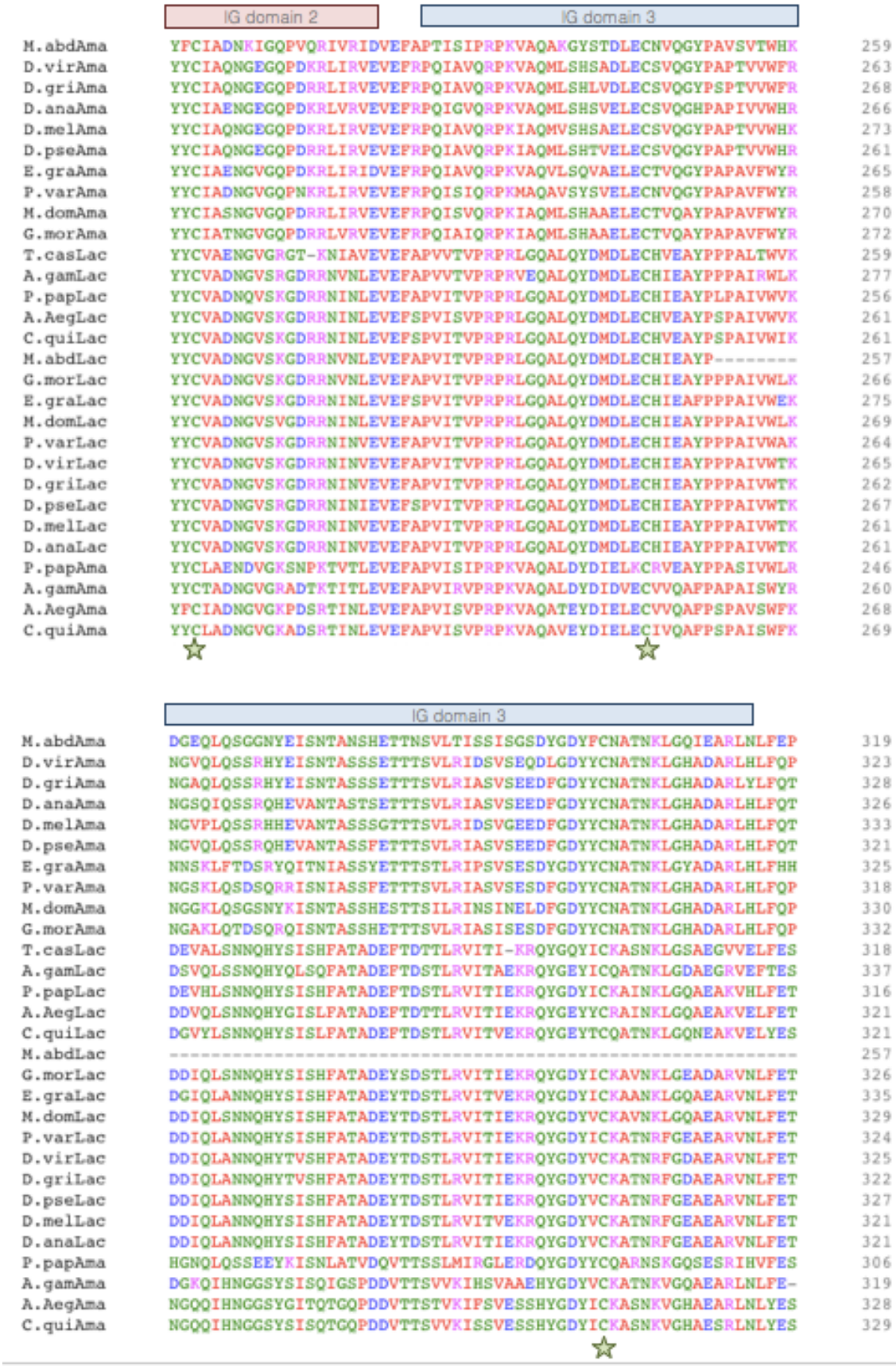
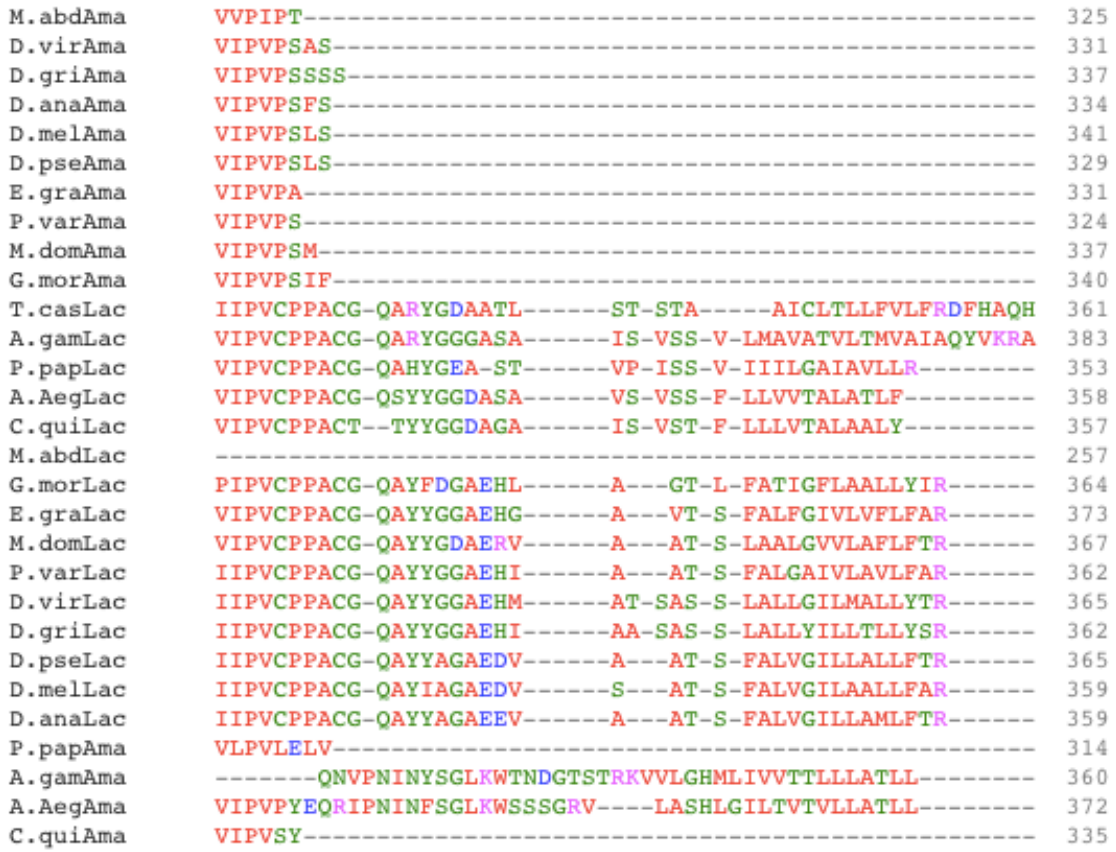


Figure 7. Phylogenetic tree of Ama and Lac orthologs
Evolutionary relationship between orthologs of Lac and Ama from 15 different insect species. 14 species are from the Dipteran order and contain both Lac and Ama. The outgroup, *Tribolium*, only contains Lac, the more ancestral gene. Red branches show the *Drosophila* species, the most derived genus in this group. Branch lengths indicate farther evolutionary divergence.


IG domain 1
M.abdAma ----------ISLLCVLAISVS-CVLGGPKITSITEETIVSEGDSVDFN 38
D.virAma --MK-SRCPRPLQL-YLW--L------LCI--SLAAASAPVISHISKDVVASVGDSVEFN 46
D.griAma --MRRMRKSDLQLV-FLW--SLLL--LSGI--SLTAATAPVISHISQDVVASVGDSVEFN 51
D.anaAma ----------MANLRLII--GLIL--ALAISLEP-ALGTPVISHISKDVVASVNDTVEFN 45
D.melAma --MQYNKRPDMARLRLLI--GLIF--CLAISLDS-VLSAPVISQISKDVVASVGDSVEFN 53
D.pseAma ----------MNNLRLRL--RLCL--CLCVGLAAAASGAPVISHITKDVVASVGDSVEFN 46
E.graAma -----MEHLKIIM-FVL-----I-----SITFRNIYAASPTISGISKDIVVSIGDTVEFN 44
P.varAma --MKLLQFGALLL-ALPW--F----------FCVGAAPPVISDITKDIVASVGDDVEFN 44
M.domAma ------MHFSQIF-LVLW--CGLSLRSRVLAASAAVSSAPVISDITPDLVASVGDDIEFN 51
G.morAma --MQLIYNLKLLF-VILW--AHLIQNA----LTAITSTSPAISGITKDIVANIGEDIEFN 51
T.casLac ------------MASV-QTTVLLVLG-VFIEICLAQRSPTISYITQTQIKDIGGTVELS 45
A.gamLac --MR-KPRSNLSTMAGLGVVLLVGIIG-GLVGPARAQRTPTISYITQEQIKDIGGTVELE 56
P.papLac ----------------MNIIAVVFVVL-PI--VCLGQRTPSISYITQEQIKDIGGTVELQ 41
A.AegLac --MS-R--------AKLNAFALVGVLL-PL--VCVAQRTPSISYITQEQIKDIGGTVELE 46
C.quiLac --MN-S--------LRVEMLVLVGVLW-PL--VVMAQRTPSISYITQEQIKDIGGTVEME 46
M.abdLac --MHLK-----SVINSRFSSSVALLFT-ILPTLSFALRTPTISYITQEQIKDIGGTVELK 52
G.morLac ---MNVSI--SSNINLWYTACILLLSL-LANKSTHAQRTPTISYISQEQIKDIGGTVEFE 54
E.graLac --MFCS----NSNMYSIG--AAVVLLC-AVISSTQAQRTPTISYITQEQIKDIGGTVEFD 51
M.domLac MSPFRSRIVKQISSRIFG--LLVTVLL-FSGVVVQAQRTPTISYITQEQIKDIGGTVEFD 57
P.varLac --MWRRYLNNISVL---L--LITVFVL-QLATHAQAQRTPTISYITQEQIKDIGGTVEFD 52
D.virLac --MWQAEHLKI--NNLWH--ALLLLLI-ALATHAQAQRTPTISYITQEQIKDIGGTVEFD 53
D.griLac --MWLKIK-----IDLWA--ALLLLLV-ALATHAEAQRTPTISYITQEQIKDIGGTVEFD 50
D.pseLac --MWRPKMVNNVGSTLAG--SLLLGLA-LLATQAEAQRTPTISYITQEQIKDIGGTVEFD 55
D.melLac --MWRPSISNCVWSTLL--------LA-IFVQQTLAQRTPTISYITQEQIKDIGGTVEFD 49
D.anaLac --MWRPSILKCVWSTLL--------LA-LACHQASGQRTPTISYITQEQIKDIGGTVEFD 49
P.papAma --MP------ESCLKFI-IFLLFAVHL-------GASQKARITDISLKQIKEIGESTHLN 44
A.gamAma ----------------MCLLPTFLPAANAQQASTASPSIVFISPEQIKDIGEQVTLN 41
A.AegAma --MG------LLRLNPL-AICLILVAANLCAA-QQQAP-PTITSISPEQIKDIGESVTFE 49
C.quiAma --MR------LLQLDLL-IICLIAVHLCKCHA-QQQPPFPTIAAISPEQIKDIGESVTLE 50
IG domain 1
M.abdAma CTVKDHGTRSLEWKKTNIDKRDDSEVLLA----------LRDTLNFADPRYTVKIIP- 85
D.virAma CTVEHVGQLTVSWAKTEPN----SAVVLS----------MRNMLSLPDQRYNISVQEN 90
D.griAma CTVEHVGQLTVSWAKTDPN----SAVVLS----------MRNMLSLPDQRYNVSVHEN 95
D.anaAma CTVEQVGQLAVSWYKLPSGPKDQAPMALS----------LRNVLSLLDERYNLTVTEG 93
D.melAma CTVEEVGQLSVSWAKRPSESD-TNSVVLS----------MRNILSLPDQRYNVTVTEG 100
D.pseAma CTVEQVGQLSVSWAKSDT------NSMLS----------MRNMLSLPDQRYNVSVTED 88
E.graAma CTVEDVGRMSVSWAKRTSETG-SGSVVLS----------MRNILSLSDPRYSILEQKD 91
P.varAma CTVENVGRMTVSWA--------LGSVVLS----------MRSILSLSDPRYTITEAKD 84
M.domAma CTVENVGRMSVSWAKRTSE---IGSVVLS----------MRNILSLSDPRYTIVETND 96
G.morAma CTVQNVGRLSVSWAKRTPKSG-SGSVVLS----------MRSILSLSDPRYTIVETKD 98
T.casLac CSVQYTNDYPVIWMKIDPRSN--AN----------SLPISTGSTLILHESRYSLRYDE- 91
A.gamLac CSVLYASDYSVHWVKTSNDRS--DTVFLS----------TGSSLVLKDSRFSLRYDL- 101
P.papLac CSVQYAKEYAVVWTKTGRDRS--DSVFLS----------TGSSLVIKDSRFAIRHDI- 86
A.AegLac CSVQYAKEYSVHWIKTGRDRS--DVVFLS----------TGSALVLKDSRFSLRYDP- 91
C.quiLac CSVQYAKEYSVHWIKTGRDRS--DVVFLS----------TGSALVLKDSRFALRFDP- 91
M.abdLac CAVQFAEEYSVNWVKIGSYG----TVFLS----------TGSTMVLKDSRFSLRHDQ- 95
G.morLac CSVQYAKEYSVQWTKTDGE-----PVFLS----------SGSTLVIRDSRFALRYEP- 96
E.graLac CSVQFAKDYSVFWLKTDND-----AVFLSTXXXXXXXXXXSTGSTLVIKDSRFALRYDP- 105
M.domLac CSVQYATDYPVIWTKNAGD-----SVFLS----------TGSTLVIKDSRFALRYDP- 99
P.varLac CSVQYAKEYSVLWTKTDTD-----PVFLS----------TGSTLVIKDSRFSLRYDP- 94
D.virLac CSVQYAKEYNVLFLKTDSD-----AVFLS----------TGSTLVIKDSRFSLRYDP- 95
D.griLac CSVQYAKEYNVLFLKTDSD-----PVFLS----------TGSTLVIKDSRFSLRYDP- 92
D.pseLac CSVQYAKEYNVLFLKTDSD-----PVFLS----------TGSTLVIKDSRFSLRYDP- 97
D.melLac CSVQYAKEYNVLFLKTDSD-----PVFLS----------TGSTLVIKDSRFSLRYDP- 91
D.anaLac CSVQYAKEYNVLFLKTDSD-----PVFLS----------TGSTLVIKDSRFSLRYDP- 91
P.papAma CTAKDVGMSKVSWNKKKLDLT--GD----------PIPISFDDQLTINDPRFKIVHDQ- 90
A.gamAma CSIANVKNYIVGWQKSNRDRT--KE----------SNIISLGVQLAVTEDRFRLNFTK- 87
A.AegAma CDIDNVGKFTVGWQKTNRERS--QN----------VNSISLGPTLAVAEDRFKVSVEK- 95
C.quiAma CDITNVGKFTVGWQKTNRERS--QQ----------VNTISLGPTLAVAEERFRVDVDK- 96


IG domain 1
IG domain 2
M.abdAma EKDGVVIYHFNIKGIEASDMGHYVCSINLSQKDKITSSVNLYVKHSPIILETKTLKTHVV 145
D.virAma TSQDSAVYSFRIRQIEASDMGPYECQVIVSASGKVTKKLNLLIKTPPVIS-ESTPKSTLV 149
D.griAma AGKDSAVYSFRIRQIEASDMGPYECQVIVSATGKVTKKLNLLIKTPPVIS-ELTPKSMLV 154
D.anaAma PAAGSATYTLRIQKVEANDGGPYECQVIVSVNEKVTKKLNLQIKTPPIIA-ETTPKSTLV 152
D.melAma PKTGSAIYTFRIQNIEVSDMGPYECQVLVSATEKVTKKLSLQIKTPPVIA-ENTPKSTLV 159
D.pseAma PKAGNAIYTFRIKQIEVNDMGPYECQVLVSTSEKITKKLNLQIKTPPVIS-ERTPKTALV 147
E.graAma NKTGSATYTFTITNIESSDMGAYECQVIVSAADKVTKKLHLSVKHPPIISEERTPKSMLV 151
P.varAma DATAKATYSFKIKKLEPSDSGYYECQVIVSSTDKVTKKVNLSIKTPPVISEERTPKSTLV 144
M.domAma DKANKATYTFKITKIEATDMGAYECQVILTATDKITKKLNLAIKHPAIISEENTPKSMLV 156
G.morAma NKTDSAVYTFKITKIETSDTGSYECQVIVAATEKITKTLNLAVKHPAIISQERTPKSMVV 158
T.casLac ---ASSTYTLQIKDIQETDAGFYHCQIIISPSNRVSAEVELQVRRPPFI-DNS-TRSVVV 146
A.gamLac ---SSTSYTLQIKDIQETDAGIYQCQVVLSVTNKISAEVALNVRRPPIISDNS-TQSLVV 157
P.papLac ---SSSTYTLQIKDIQETDAGIYQCQVILSVVNKITADVELQVRRPPVISDNS-TQSIVA 142
A.aegLac ---SSSSYILQVKDIQETDAGIYQCQVVLSVTNKITADVELQVRRPPIISDNS-TQSLVV 147
C.quiLac ---SSSSYILQIKDIQETDAGIYQCQVVLSVTNKITADVELQVRRPPIISDNS-TQSLVA 147
M.abdLac ---ASTTYTLQIKDLQETDAGKYQCEVVLSVSNKVTGEVLLSVRRPPIISDNS-TQSIVA 151
G.morLac ---SASTYRLQIKDIQETDAGTYTCQVVISVVNKVSANVKLSVRRPPVISDNS-TQSLVA 152
E.graLac ---NSSTYKLQIKDIQETDAGIYTCQVVISVVSKVSADVKLSVRRPPVISDNS-TQSLVA 161
M.domLac ---NSSTYKLQIKDIQETDAGQYTCQIVMSVTHKVSANVMLSVRRPPVISDNS-TQSLVA 155
P.varLac ---NSSTYKLQIKDIQETDAGTYTCQVVISVVHKVTANVKLSVRRPPVISDNS-TQSIVA 150
D.virLac ---NSSTYKLQIKDIQETDAGTYTCQVVISTVHKVSASVKLSVRRPPVISDNS-TQSIVA 151
D.griLac ---NSSTYKLQIKDIQETDAGTYTCQVVISTVHKVSANVKLSVRRPPVISDNS-TQSIVA 148
D.pseLac ---SSSTYKLQIKDIQETDAGTYTCQVVISTVHKVSAEVKLSVRRPPVISDNS-TQSIVA 153
D.melLac ---NSSTYKLQIKDIQETDAGTYTCQVVISTVHKVSAEVKLSVRRPPVISDNS-TQSVVA 147
D.anaLac ---SSSTYKLQIKDIQETDAGTYTCQVVISVVHKVSADVKLSVRRPPVISDNS-TQSIVA 147
P.papAma -----DTYSLQISNIGSTDMGIYECQIPVSFDERPTSSVELLVRHPPVISEKLSTRVAQV 145
A.gamAma -ENNAANYVLEIHDIVNTDAGLYECQIQVNSTSKITKTVELQVRHPPMLLENQHTNTLTK 146
A.aegAma -ENNTMNYTLTISDIVNTDAGLYECQIQVNSTNKVTATVELQVRHPPLLQDNLMATTVTK 154
C.quiAma -ENNTMKYKLTISDIINTDAGLYECQIQVNSTSKVTNTVELQVRHPPILQDNLMSTTVTK 155
IG domain 2
M.abdAma TEGYKLEALCQADGYPKPTISWKRENNAIMPGG------GQVFNGNTLKIKETHRLDRGN 199
D.virAma TEGQNLEVSCHANGFPTPTISWAREQNAIMPAG------GHVLNEPTLRIKTVHRLDRGG 203
D.griAma TEGQNLEISCHASGFPAPTISWAREQNAIMPAG------GHLLNEPTLRIRAIHRVDRGG 208
D.anaAma TEGQNLELTCHANGFPKPTISWAREHNGVMPAG------GHLLSEPTLRIKSVHRLDRGG 206
D.melAma TEGQNLELTCHANGFPKPTISWAREHNAVMPAG------GHLLAEPTLRIRSVHRMDRGG 213
D.pseAma TEGENLELTCHANGFPKPTIAWARENNAIMPAG------GHLLAEPTLRIRTVHRMDRGG 201
E.graAma TEGENLEVNCHANGFPQPTISWARADNSIMPAG------GYVQHGPSLRIKQVHRLDRGG 205
P.varAma TEGQNLEVTCHANGFPQPTISWKREGNAIMPAG------GHTLAGPTLRIKEAHRLDRGG 198
M.domAma TEGQNLEISCHADGFPQPTISWERENRAIMPAG------GSSFDGQTLRIKEAHRLDRGG 210
G.morAma TEGESLEITCYADGFPQPTLSWERAHNAIMPAG------GQMHLGATLRIKEVHRLDRGA 212
T.casLac SEGQAVQMECYAGGYPPPRISWRRENNAILPTG------GSIYRGNVMKIKQIKKEDRGT 200
A.gamLac SEGQPAQMECYASGYPVPQITWRRENNAILPTGNDGGGGGATYSGNVLNIHSVHKEDRGT 217
P.papLac SEGKSVQMECYAAGYPRPTITWRRENNAILPTG------GSTYTGNVLKINSVRKEDRGT 196
A.aegLac SEGQSVQMECYASGYPPPQITWRRENNAILPTG------GAIYTGNVMKINSVQKEDRGT 201
C.quiLac SEGEAVMMECYASGYPPPQITWRRENNAILPTG------GAIYTGNVMKINSVRKEDRGT 201
M.abdLac TEGQAVQMECFAGGYPIPTITWRRENNAILPTD------SATYVGNILKINSVKKEDRGT 205
G.morLac SEGSEVQMECYASGYPTPTITWRRENNAILPTD------SATFVGNILRIKSVKKEDRGT 206
E.graLac TEGAEVEMNCYASGYPTPTITWRRENNAILPTD------SATYVGNVLKIKSVKKEDRGT 215
M.domLac SEGSEVQMECYASGYPTPTISWRRENNAILPTD------SATYVGNILKIKSVKKEDRGT 209
P.varLac NEGGEVQMECYASGYPTPTITWRRENNAILPTD------SATYTGNILRIKSVRKEDRGT 204
D.virLac SEGSEVQMECYASGYPTPTITWRRENNAILPTD------SATYVGNILRIKSVKKEDRGT 205
D.griLac SEGSEVQMECYASGYPTPSITWRRENNAILPTD------SATYVGNILRIKSVKKEDRGT 202
D.pseLac SEGSEVQMECFASGYPTPTITWRRENNAILPTD------SATYVGNILRIKSVKKEDRGT 207
D.melLac SEGSEVQMECYASGYPTPTITWRRENNAILPTD------SATYVGNTLRIKSVKKEDRGT 201
D.anaLac SEGTEVQMECYASGYPTPTITWRRENNAILPTD------SATYVGNILRIKSVKKEDRGT 201
P.papAma AEHQPVNLECYADGYPRPSITWTRD----------------GNVLKITSTRKEDRGT 186
A.gamAma AEGEDAQLVCRAEGYPRPTISWRREYNAILPIG------GQTFTGNELRLNGLRREDRGT 200
A.aegAma AEGENVKLTCSAEGYPRPTISWKREYGAILPIG------GQSYTGNELSLSSLVREDRGT 208
C.quiAma AEGEDVKLSCVAEGYPRPSITWKREYNAILPIG------GHSFSGNELSLSSLAKEDRGP 209


IG domain 2
IG domain 3
M.abdAma YFCIADNKIGQPVQRIVRIDVEFAPTISIPRPKVAQAKGYSTDLECNVQGYPAVSVTWHK 259
D.virAma YYCIAQNGEGQPDKRLIRVEVEFRPQIAVQRPKVAQMLSHSADLECSVQGYPAPTVVWFR 263
D.griAma YYCIAQNGEGQPDRRLIRVEVEFRPQIAVQRPKVAQMLSHLVDLECSVQGYPSPTVVWFR 268
D.anaAma YYCIAENGEGQPDKRLVRVEVEFRPQIGVQRPKVAQMLSHSVELECSVQGHPAPIVVWHR 266
D.melAma YYCIAQNGEGQPDKRLIRVEVEFRPQIAVQRPKIAQMVSHSAELECSVQGYPAPTVVWHK 273
D.pseAma YYCIAQNGEGQPDRRLIRVEVEFRPQIAVQRPKIAQMLSHTVELECSVQGYPAPTVVWHR 261
E.graAma YYCIAENGVGQPDKRLIRIDVEFRPQIAVQRPKVAQVLSQVAELECTVQGYPAPAVFWYR 265
P.varAma YYCIADNGVGQPNKRLIRVEVEFRPQISIQRPKMAQAVSYSVELECNVQGYPAPAVFWYR 258
M.domAma YYCIASNGVGQPDRRLIRVEVEFRPQISVQRPKIAQMLSHAAELECTVQAYPAPAVFWYR 270
G.morAma YYCIATNGVGQPDRRLVRVEVEFRPQIAIQRPKIAQMLSHAAELECTVQAYPAPAVFWYR 272
T.casLac YYCVAENGVGRGT-KNIAVEVEFAPVVTVPRPRLGQALQYDMDLECHVEAYPPPALTWVK 259
A.gamLac YYCVADNGVSRGDRRNVNLEVEFAPVVTVPRPRVEQALQYDMDLECHIEAYPPPAIRWLK 277
P.papLac YYCVADNQVSKGDRRNINLEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPLPAIVWVK 256
A.aegLac YYCVADNGVSKGDRRNINLEVEFSPVISVPRPRLGQALQYDMDLECHVEAYPSPAIVWVK 261
C.quiLac YYCVADNGVSKGDRRNINLEVEFSPVITVPRPRLGQALQYDMDLECHVEAYPSPAIVWIK 261
M.abdLac YYCVADNGVSKGDRRNVNLEVEFAPVITVPRPRLGQALQYDMDLECHIEAYP-------- 257
G.morLac YYCVADNGVSKGDRRNVNLEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWLK 266
E.graLac YYCVADNGVSKGDRRNINLEVEFSPVITVPRPRLGQALQYDMDLECHIEAFPPPAIVWEK 275
M.domLac YYCVADNGVSVGDRRNINLEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWLK 269
P.varLac YYCVADNGVSKGDRRNINVEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWAK 264
D.virLac YYCVADNGVSKGDRRNINVEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWTK 265
D.griLac YYCVADNGVSKGDRRNINVEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWTK 262
D.pseLac YYCVADNGVSRGDRRNINIEVEFSPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWTK 267
D.melLac YYCVADNGVSKGDRRNINVEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWTK 261
D.anaLac YYCVADNGVSKGDRRNINVEVEFAPVITVPRPRLGQALQYDMDLECHIEAYPPPAIVWTK 261
P.papAma YYCLAENDVGKSNPKTVTLEVEFAPVISIPRPKVAQALDYDIELKCRVEAYPPASIVWLR 246
A.gamAma YYCTADNGVGRADTKTITLEVEFAPVIRVPRPKVAQALDYDIDVECVVQAFPAPAISWYR 260
A.aegAma YFCIADNGVGKPDSRTINLEVEFAPVISVPRPKVAQATEYDIELECVVQAFPSPAVSWFK 268
C.quiAma YYCLADNGVGKADSRTINLEVEFAPVISVPRPKVAQAVEYDIELECIVQAFPSPAISWFK 269
IG domain 3
M.abdAma DGEQLQSGGNYEISNTANSHETTNSVLTISSISGSDYGDYFCNATNKLGQIEARLNLFEP 319
D.virAma NGVQLQSSRHYEISNTASSSETTTSVLRIDSVSEQDLGDYYCNATNKLGHADARLHLFQP 323
D.griAma NGAQLQSSRHYEISNTASSSETTTSVLRIASVSEEDFGDYYCNATNKLGHADARLYLFQT 328
D.anaAma NGSQIQSSRQHEVANTASTSETTTSVLRIASVSEEDFGDYYCNATNKLGHADARLHLFQT 326
D.melAma NGVPLQSSRHHEVANTASSSGTTTSVLRIDSVGEEDFGDYYCNATNKLGHADARLHLFQT 333
D.pseAma NGVQLQSSRQHEVANTASSFETTTSVLRIASVSEEDFGDYYCNATNKLGHADARLHLFQT 321
E.graAma NNSKLFTDSRYQITNIASSYETTTSTLRIPSVSESDYGDYYCNATNKLGYADARLHLFHH 325
P.varAma NGSKLQSDSQRRISNIASSFETTTSVLRIASVSESDFGDYYCNATNKLGHADARLHLFQP 318
M.domAma NGGKLQSGSNYKISNTASSHESTTSILRINSINELDFGDYYCNATNKLGHADARLHLFQP 330
G.morAma NGAKLQTDSQRQISNTASSHETTTSVLRIASISESDFGDYYCNATNKLGHADARLHLFQP 332
T.casLac DEVALSNNQHYSISHFATADEFTDTTLRVITI-KRQYGQYICKASNKLGSAEGVVELFES 318
A.gamLac DSVQLSSNQHYQLSQFATADEFTDSTLRVITAEKRQYGEYICQATNKLGDAEGRVEFTES 337
P.papLac DEVHLSNNQHYSISHFATADEFTDSTLRVITIEKRQYGDYICKAINKLGQAEAKVHLFET 316
A.aegLac DDVQLSNNQHYGISLFATADEFTDTTLRVITIEKRQYGEYYCRAINKLGQAEAKVELFET 321
C.quiLac DGVYLSNNQHYSISLFATADEFTDSTLRVITVEKRQYGEYTCQATNKLGQNEAKVELYES 321
M.abdLac ---------- 257
G.morLac DDIQLSNNQHYSISHFATADEYSDSTLRVITIEKRQYGDYICKAVNKLGEADARVNLFET 326
E.graLac DGIQLANNQHYSISHFATADEYTDSTLRVITVEKRQYGDYICKAANKLGQAEARVNLFET 335
M.domLac DDIQLSNNQHYSISHFATADEYTDSTLRVITIEKRQYGDYVCKAVNKLGQAEARVNLFET 329
P.varLac DDIQLANNQHYSISHFATADEYTDSTLRVITIEKRQYGDYICKATNRFGEAEARVNLFET 324
D.virLac DDIQLANNQHYTVSHFATADEYTDSTLRVITIEKRQYGDYVCKATNRFGDAEARVNLFET 325
D.griLac DDIQLANNQHYTVSHFATADEYTDSTLRVITIEKRQYGDYVCKATNRFGDAEARVNLFET 322
D.pseLac DDIQLANNQHYSISHFATADEYTDSTLRVITIEKRQYGDYVCKATNRFGEAEARVNLFET 327
D.melLac DDIQLANNQHYSISHFATADEYTDSTLRVITVEKRQYGDYVCKATNRFGEAEARVNLFET 321
D.anaLac DDIQLANNQHYSISHFATADEYTDSTLRVITIEKRQYGDYVCKATNRFGEAEARVNLFET 321
P.papAma HGNQLQSSEEYKISNLATVDQVTTSSLMIRGLERDQYGDYYCQARNSKGQSESRIHVFES 306
A.gamAma DGKQIHNGGSYSISQIGSPDDVTTSVVKIHSVAAEHYGDYVCKATNKVGQAEARLNLFE- 319
A.aegAma NGQQIHNGGSYGITQTGQPDDVTTSTVKIPSVESSHYGDYICKASNKVGHAEARLNLYES 328
C.quiAma NGQQIHNGGSYSISQTGQPDDVTTSVVKISSVESSHYGDYICKASNKVGHAESRLNLYES 329



Supplementary Figure 1. Multiple sequence alignment of Ama and Lac orthologs
Multiple sequence alignment of Lac and Ama translated orthologs from 15 different insect species. The three IG domains are shown, and stars mark where the two conserved cysteines are located within each domain.

REFERENCES

Barthalay, Y., Hipeau-Jacquottel, R., De La Escalera, S., Jimenez, F., & Piovant, M. (1990). Drosophila neurotactin mediates heterophilic cell adhesion. *The EMBO Journal*, *9*(1), 3603–3609.

Darboux, I., Barthalay, Y., Piovant, M., & Hipeau-Jacquotte, R. (1996). The structure-function relationships in Drosophila neurotactin show that cholinesterasic domains may have adhesive properties. *The EMBO Journal*, *15*(18), 4835–4843.

de la Escalera, S., Bockamp, E. O., Moya, F., Piovant, M., & Jiménez, F. (1990). Characterization and gene cloning of neurotactin, a Drosophila transmembrane protein related to cholinesterases. *The EMBO Journal*, *9*(11), 3593–601.

Frémion, F., Darboux, I., Diano, M., Hipeau-Jacquotte, R., Seeger, M. A., & Piovant, M. (2000). Amalgam is a ligand for the transmembrane receptor neurotactin and is required for neurotactin-mediated cell adhesion and axon fasciculation in Drosophila. *The EMBO Journal*, *19*(17), 4463–72.

Karlstrom, R. O., Wilder, L. P., & Bastiani, M. J. (1993). Lachesin: an immunoglobulin superfamily protein whose expression correlates with neurogenesis in grasshopper embryos. *Development (Cambridge, England)*, *118*(2), 509–522.

Liebl, E. C., Rowe, R. G., Forsthoefel, D. J., Stammler, A. L., Schmidt, E. R., Turski, M., & Seeger, M. A. (2003). Interactions between the secreted protein Amalgam, its transmembrane receptor Neurotactin and the Abelson tyrosine kinase affect axon pathfinding. *Development (Cambridge, England)*, *130*(14), 3217–3226.

Llimargas, M. (2004). Lachesin is a component of a septate junction-based mechanism that controls tube size and epithelial integrity in the Drosophila tracheal system. *Development*, *131*(1), 181–190.

Seeger, M. A., Haffley, L., & Kaufman, T. C. (1988). Characterization of amalgam: A Member of the Immunoglobulin Superfamily from Drosophila. *Cell*, *55*, 589–600.